

Lecture 2: 知识 - 计算机性能指标

存储衡量单位

概念：字、字长、机器字长、指令字长、存储字长

- 字长：
 - 是指 CPU 内部用于整数运算的数据通路的宽度；
 - 因此字长等于 CPU 内部用于整数运算的ALU位数和通用寄存器宽度，它反映了计算机处理信息的能力；
- 字：
 - 字用来表示被处理信息的单位，用来度量数据类型的宽度；
 - 举例：
 - 如 x86 机器中将一个字定义为 16 位；
- 指令字长：
 - 一个指令字中包含的二进制代码的位数；
 - 必须是字节的整数倍；
 - 指令字长一般取存储字长的整数倍，若指令字长等于存储字长的 2 倍，则需要 2 个访存周期来取出一条指令；若指令字长等于存储字长，则取指令周期等于机器周期；
- 存储字长：
 - 一个存储单元存储的二进制代码的位数；
 - 必须是字节的整数倍
- 机器字长：
 - 机器字长通常与 CPU 的寄存器位数、ALU 有关，因此字长一般等于通用寄存器的位数或 ALU 的宽度；

概念：数据通路带宽

- 指数据总线一次所能并行传送信息的位数；
- 这里所说的数据通路宽度是指外部数据总线的宽度，它与 CPU 内部的数据总线宽度（内部寄存器的大小）有可能不同；

概念：主存容量

- 概念：
 - 指主存储器所能存储信息的最大容量，通常以字节来衡量，也可用 字数 x 字长（如 512 K * 16 位）来表示存储容量；
- MAR 与 MDR：
 - MAR 的位数反映了存储单元的个数；
 - MDR 的位数反映了存储单元的字长；
- 举例：
 - MAR 为 16 位，表示 $2^{16} = 65536$ ，即此存储体内有 65536 个存储单元（可称为 64K 内存，1K=1024）

- MDR 为 32 位，则表示存储容量为 $64K * 32$ 位

运算速度指标

概念：吞吐量和响应时间

- 吞吐量：
 - 指系统在单位时间内**处理请求的数量**；
 - 它取决于信息能多快地输入内存，CPU 能多快地取指令，数据能多快地从内存取出或存入，以及所得结果能多快地从内存送给一台外部设备；
 - 几乎每步都关系到主存储器；
 - 系统吞吐量主要**取决于主存储器的存取周期**；
- 响应时间：
 - 指从用户向计算机发送一个请求，到系统对该请求做出响应并获得所需结果的**等待时间**；
 - 包括：
 - (1) CPU 时间 (运行一个程序所花费的时间)；
 - (2) 等待时间 (用于磁盘访问、存储器访问、I/O 操作、操作系统开销等的时间)；

概念：主频与时钟周期

- 时钟脉冲信号：
 - 由机器脉冲源发出的脉冲信号经整形和分频后形成；
- 时钟周期：
 - 概念：
 - 机器内部**主时钟脉冲信号的宽度**，它是 CPU 工作的**最小时间单位**；
 - 时钟周期的限制因素：
 - 时钟周期以相邻状态单元间**组合逻辑电路的最大延迟**为基准确定；
 - 时钟周期也以指令流水线的**每个流水段的最大延迟时间**确定；
- 主频：
 - 概念：
 - 又称 CPU 时钟频率
 - 表示机器内部**主时钟的频率**，即**时钟周期的倒数**；
 - 主频最直观的理解就是**每秒有多少个时钟周期**；
 - 意义：
 - 衡量机器速度的重要参数；
 - 对于同一个型号的计算机，其主频越高，完成指令的一个执行步骤所用的时间越短，执行指令的速度越快；

概念：CPI

- 概念：
 - CPI : Cycle Per Instruction
 - 指的是执行**一条指令**所需时钟周期数
 - CPI 是一个平均值；

- 不同指令的时钟周期数可能不同 -> 对于一个程序或一台机器来说，其 **CPI** 指该程序或该机器指令集中的**所有指令执行所需的平均时钟周期数**；
- 分析：不同处理器类型下的 **CPI**
 - 真题：2020年
 - (1) 单周期 CPU：CPI=1
 - 单周期 CPU 中所有指令的指令周期为一个时钟，因此每条指令的 **CPI** 为 1；
 - (2) 多周期 CPU：CPI>1
 - 多周期 CPU 的 **CPI>1**，因为在多周期 CPU 中，一条指令的执行需要几个时钟周期完成，每个时钟周期对应执行过程的一个阶段；
 - (3) 基本流水线：
 - 一般情况：CPI>1
 - 理想情况：CPI=1
 - 对于基本流水线，在理想情况下，其每个时钟周期均留出一条指令，因此 CPI=1；
 - (4) 超标量流水线：CPI<1
 - 在每个时钟周期内，超标量流水线 **CPI** 会并行执行多条流水线，CPI<1；

概念：IPS

- 概念：
 - **IPS**：Instruction Per Second
 - 每秒执行多少条**指令**
- 公式：

$$IPS = \frac{\text{主频}}{\text{平均CPI}}$$

MIPS

- 概念：
 - 每秒执行多少百万条指令；
- 公式：

$$MIPS = \text{指令条数} / (\text{执行时间} * 10^6) = \text{主频} \div (CPI * 10^6)$$

FLOPS

- 概念：
 - **Floating-point Operations Per Second** 每秒执行多少次浮点运算
- MFLOPS (Million FLOPS)，即每秒执行多少百万 (10^6) 次浮点运算。
- GFLOPS (Giga FLOPS)，即每秒执行多少十亿 (10^9) 次浮点运算。
- TFLOPS (Tera FLOPS)，即每秒执行多少万亿 (10^{12}) 次浮点运算。
- PFLOPS (Peta FLOPS)，即每秒执行多少千万亿 (10^{15}) 次浮点运算。
- EFLOPS (Exa FLOPS)，即每秒执行多少百京 (10^{18}) 次浮点运算 (1 京 = 1 亿亿 = 10^{16})。
- ZFLOPS (Zetta FLOPS)，即每秒执行多少十万京 (10^{21}) 次浮点运算。
- 总结：
 - **K/M/G/T/P/E/Z** 每个相隔三次方

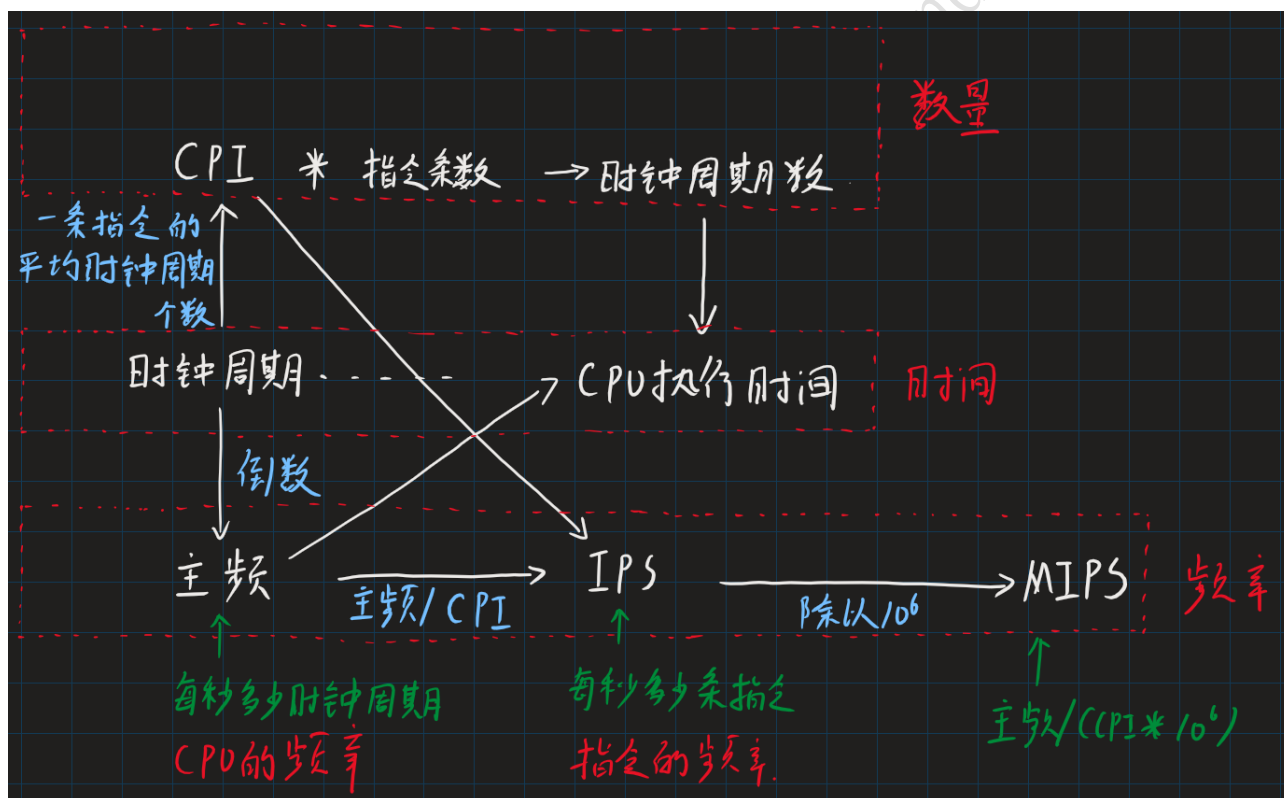
- 1 亿 = 10^8
- 举例:
 - $93\text{PFLOPS} = 93 * 10^{15} = 9.3 * 10^{16} = 9.3 \text{ 亿亿}$

概念: CPU 执行时间

- 概念:
 - CPU 的性能取决于三个要素: 主频、CPI 和指令条数;
- 公式:
 - 指令条数 * CPI = 这么多条指令共需要的时钟周期数

$$\text{CPU 执行时间} = \frac{\text{CPU 时钟周期数}}{\text{主频}} = \frac{\text{指令条数} * \text{CPI}}{\text{主频}}$$

总结: 运行效率指标总结



- 注意: k 和 K 的区别

速率即数据率或称数据传输率或比特率。

比特 1/0 位

连接在计算机网络上的主机在数字信道上传送数据位数的速率。

单位是b/s, kb/s, Mb/s, Gb/s, Tb/s



速率

千 $1\text{kb/s}=10^3\text{b/s}$

兆 $1\text{Mb/s}=10^3\text{kb/s}=10^6\text{b/s}$

吉 $1\text{Gb/s}=10^3\text{Mb/s}=10^6\text{kb/s}=10^9\text{b/s}$

太 $1\text{Tb/s}=10^3\text{Gb/s}=10^6\text{Mb/s}=10^9\text{kb/s}=10^{12}\text{b/s}$

存储容量 1Byte (字节) = 8bit (比特)

$1\text{KB}=2^{10}\text{B}=1024\text{B}=1024 * 8\text{b}$

$1\text{MB}=2^{10}\text{KB}=1024\text{KB}$

$1\text{GB}=2^{10}\text{MB}=1024\text{MB}$

$1\text{TB}=2^{10}\text{GB}=1024\text{GB}$

获取笔记持续更新可加微信: WandD222